Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

A decade of genome sequencing has revolutionized studies of experimental evolution

Eric L Bruger and Christopher J Marx

Genome sequencing has revolutionized studies using experimental evolution of microbes because it readily provides comprehensive insight into the genetic bases of adaptation. In this perspective we discuss applications of sequencing-based technologies used to study evolution in microbes, including genomic sequencing of isolated evolved clones and mixed evolved populations, and also the use of sequencing methods to follow the fate of introduced variations, whether neutral barcodes or variants introduced by genome editing. Collectively, these sequencing-based approaches have vastly advanced the examination of evolution in the lab, as well as begun to synthesize this work with examination of the genetic bases of adaptation and evolutionary dynamics within natural populations.

Address
Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies, Center for Modeling Complex Interactions, University of Idaho, Moscow, ID 83844, USA

Corresponding author: Marx, Christopher J (cmarx@uidaho.edu)

Current Opinion in Microbiology 2018, **45**:149–155

This review comes from a themed issue on **Microbial systems biology**

Edited by **Terence Hwa** and **Uwe Sauer**

https://doi.org/10.1016/j.mib.2018.03.002

Experimental evolution before the application of genome sequencing

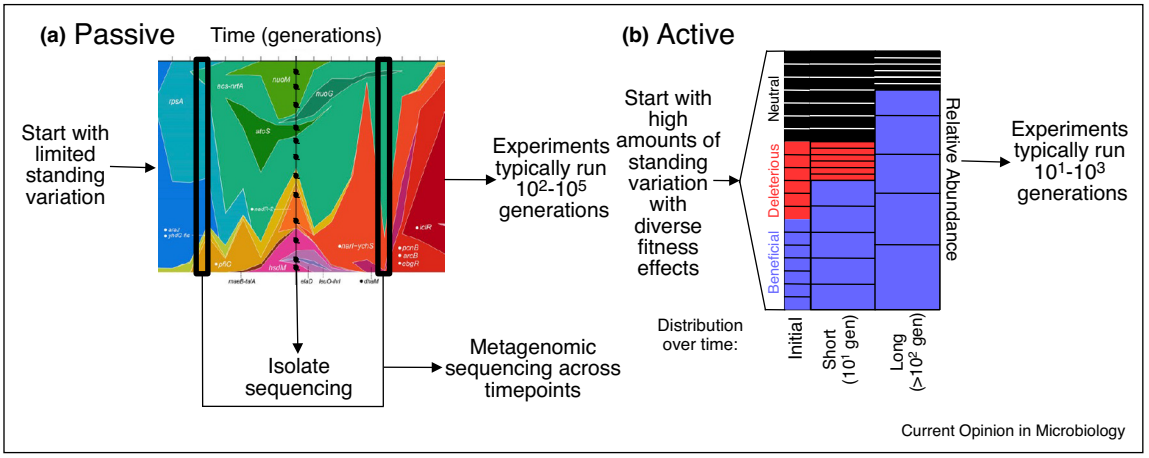
Studies delving into microbial evolution date back to early experiments involving pond microbes conducted by the reverend William Henry Dallinger in the late 1800s [1,2]. In the second half of the 20th century, pioneered by researchers such as Bruce Levin, Dan Dykhuizen, and colleagues, the use of evolution experiments in the laboratory became increasingly popular [3–8]. The attraction to this approach was the ability to precisely control the selective environment, transfer regime, and initial genotype, thereby seeding replicate populations that can be cryopreserved as a living fossil record. Upon resuscitation, comparisons could then be made through time, between lineages, and across experiments. An extensive amount was learned about changes in *phenotype* that occur during adaptation, best exemplified by a fruitful series of discoveries from Rich Lenskis long-term evolution experiment (LTEE) with *Escherichia coli* [9•,10•]. Stepping back from particulars, some commonalities have emerged from the LTEE and other similar experiments. Perhaps most prominently, the rate of adaptation is almost always fastest early in the experiment, and slows as increasing generations accumulate [10•,11]. Conversely, other phenomena were found to behave quite differently depending upon the organism and experiment in question, such as whether replicate populations would exhibit parallelism or divergence in phenotypic changes, or in the extent of tradeoffs between fitness in the selective environment versus alternative environments [12]. Unfortunately, in these early studies there was generally an inability to link these changes in phenotype with mutations that occurred to alter the *genotype* [9•,13].

Although these numerous experimental evolution studies constituted what was then called 'population genetics without the genetics' [14], in the more than a decade since the first application of whole genome sequencing to experimental evolved populations [15•] it is hard to imagine anything further from the truth. Genome sequencing and other related sequencing-based technologies have led to unprecedented progress in the study of microbial evolution in the laboratory [16•], and increasingly have been extended to studying evolution in natural environments. Here we will first discuss the purely *passive*, observational role that sequencing has played in earlier investigations following changes in experimental populations (Figure 1a). We follow this with a discussion of how sequencing can provide the key output data for experimental designs where the researcher plays an *active* role in generating variation prior to the initiation of adaptation (Figure 1b).

Sequencing individual isolates reveals evolved genotypes

The most straightforward use of genome sequencing to understand evolution is to determine the complete genome sequence of individual evolved isolates. Researchers using viruses as model systems had been using standard Sanger sequencing for this purpose much earlier [17,18], but the use of 454 sequencing to determine the genetic basis of adaptation in an experiment with *Myxococcus xanthus* [15•] was the first in a wave of papers using whole genome sequencing to uncover the genetic bases of adaptation in numerous bacterial

Figure 1



Different means for applying sequencing approaches to evolution experiments. **(a)** Passive approaches include isolate as well as metagenomic sequencing to capture information on the diversity of mutations that evolve in experimental populations. Figure adapted from [23•]. **(b)** Active approaches arise from methods that allow the generation and/or construction of large numbers of initial variants — neutral barcodes or at loci under selection — and tracking them over time. A short experimental timeframe permits observation of the various rates at which deleterious mutations are lost and neutral mutations will remain at steady frequencies, whereas a longer timeframe will see the neutral mutations begin to be squeezed out by the rising mean fitness of the population, but the relative differences in the beneficial mutational effects become more prominent.

systems. This approach provides the number, type, and targets of mutations, and it unambiguously reveals that these mutations are linked together as a genotype (Box 1). Assuming genetic manipulation is possible for the organism of interest, it is then possible to parse apart which of these mutations contribute to these phenotypes. Experiments that manipulate combinations of mutant alleles reveal both specific answers about adaptation of a particular organism to a particular environment (e.g. [19]), and illuminate general trends about adaptation, such as that beneficial mutations are generally less and less beneficial when present upon backgrounds with higher fitness (i.e., diminishing returns epistasis [20,21]).

Box 1 What to expect when you sequence evolved isolates?

Investigators new to using sequencing as part of their experimental studies are often (justifiably) curious about what they should expect to see from their experimental results. Years of isolate sequencing have provided ample information on a number of general trends that consistently crop up in evolution experiments (many of these were highlighted in [66••]), including:

- Observed biases toward non-synonymous changes selected more commonly over synonymous changes within genes.
- More mutations in promoters than expected by chance.
- A high proportion of mutations caused by insertion sequence (IS) element transposition and/or homologous recombination between multiple copies of the same IS.
- Parallelism in the loci containing beneficial mutations between replicate lineages, but generally not to the same site/SNPs. This is especially true for loss-of-function alleles that are beneficial for fitness.
- Patterns of mutations and direct allelic exchange experiments indicate an overwhelming pattern of positive selection upon beneficial mutations, with the exception of strains that become mutators. Mutators display a much wider spectrum of mutational targets and effects observed.

Whereas obtaining a single whole genome sequence for an evolved isolate was astonishing in 2006, this has become absolutely trivial at this point, and the low hurdle for sequencing has remarkably altered the types of scientific questions that can be asked. One great advantage has been the ability to sequence isolates from a previously unparalleled number of independent evolution experiments, thereby obtaining a reasonably-sized sample of what is possible for that strain placed in the selective conditions used. For example, by sequencing isolates from 120 separate populations of *E. coli* evolved to grow at an elevated temperature, it became possible to use the occurrence (or nonoccurrence) of mutations together in the same genotype more (or less) frequently than random expectation to reveal positive (or negative) epistasis – non-additive fitness effect between mutations — between them [22••] (Figure 2a). This readily revealed multiple distinct evolutionary trajectories that were possible. If the power of sequencing many isolates is instead directed at multiple isolates from multiple timepoints in a single population, it becomes possible to loosely infer clonal dynamics of these populations [23•]. Although it was once thought that beneficial alleles arise and escape drift rarely enough that they would rise in frequency and fix one at a time (i.e., periodic selection, [24]), genomic analyses of isolates (and populations, see below) have made it abundantly clear that allele dynamics in populations are exceptionally messy due to multiple lineages with beneficial mutations arising at the same time and